

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2001-051692

(43)Date of publication of application : 23.02.2001

(51)Int.Cl. G10L 15/18
G10L 15/06

(21)Application number : 11-228423 (71)Applicant : SHARP CORP
REAL WORLD COMPUTING
PARTNERSHIP

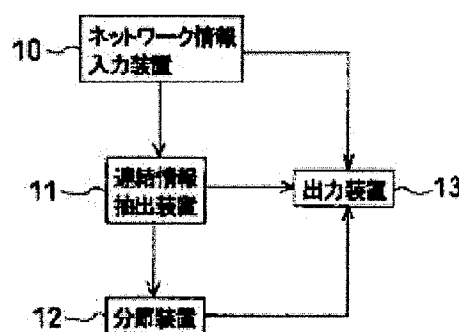
(22)Date of filing : 12.08.1999 (72)Inventor : MUKAI MICHIAKI
OKA RYUICHI

(54) AUTOMATIC SEGMENTING DEVICE FOR NETWORK AND STORAGE MEDIUM RECORDING AUTOMATIC SEGMENTING PROGRAM

(57)Abstract:

PROBLEM TO BE SOLVED: To provide an automatic segmenting device for a network automatically extracting a characteristic connection structure from the network showing a structure of a sign series.

SOLUTION: This device is provided with a network information input device 10 reading in a network structure showing the sign series, a connective information extracting device 11 extracting connective information of each node from the network structure, a segmenting device 12 segmenting the network based on the connective information and an output device 13 outputting a segmented result. By the automatic segmenting device for the network and a recording medium recording an automatic segmenting program, the connective structure of the network showing the structure of the sign series is analyzed, and is segmented based on the characteristics of the connective structure. Since this processing is performed from an initial state, a user doesn't have to perform initial work by manual operation, and extracts automatically a characteristic part.



(19)日本国特許庁(JP)

(12) 公開特許公報(A)

(11)特許出願公開番号

特開2001-51692

(P2001-51692A)

(43)公開日 平成13年2月23日(2001.2.23)

(51)Int.Cl.⁷

識別記号

FI

テーマコード(参考)

G10L 15/18

G10L 3/00

537C 5D015

15/06

521J

審査請求 未請求 請求項の数3 OL (全6頁)

(21)出願番号 特願平11-228423

(22)出願日 平成11年8月12日(1999.8.12)

(71)出願人 000005049

シャープ株式会社

大阪府大阪市阿倍野区長池町22番22号

(71)出願人 593162453

技術研究組合新情報処理開発機構

東京都千代田区東神田2-5-12 龍角散ビル8階

(72)発明者 向井 理朗

大阪府大阪市阿倍野区長池町22番22号 シャープ株式会社内

(74)代理人 100091096

弁理士 平木 祐輔

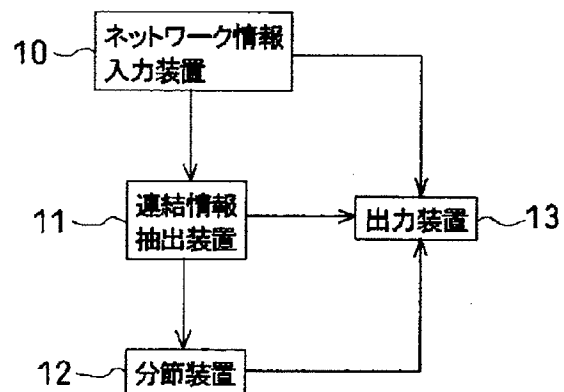
最終頁に続く

(54)【発明の名称】 ネットワークの自動分節装置及び自動分節プログラムを記録した記録媒体

(57)【要約】

【課題】 記号系列の構造を示すネットワークから特徴のある接続構造を自動抽出することができるネットワークの自動分節装置を提供する。

【解決手段】 記号系列を示すネットワーク構造を読み込むネットワーク情報入力装置10と、ネットワーク構造から各ノードの連結情報を抽出する連結情報抽出装置11と、前記連結情報を元にネットワークを分節する分節装置12と、分節した結果を出力する出力装置13とを備える。



【特許請求の範囲】

【請求項1】 記号系列の構造を示すネットワークの自動分節装置であって、ネットワーク構造を読み込む読込手段と、該読込手段によって読み込まれたネットワーク構造から各ノードの連結情報を抽出する連結情報抽出手段と、該連結情報抽出手段によって抽出された各ノードの連結情報を元にネットワークを分節する分節手段と、該分節手段によって分節された結果を出力する出力手段と、を備えることを特徴とするネットワークの自動分節装置。

【請求項2】 記号系列の構造を示すネットワークの自動分節装置であって、ネットワーク構造を読み込む読込手段と、該読込手段によって読み込まれたネットワーク構造から各ノードの連結情報を抽出する連結情報抽出手段と、該連結情報抽出手段によって抽出された各ノードの連結情報から所定のしきい値よりも頻度が少ないアークを除き、これにより入力アークが存在しなくなったノードを除いたネットワークを作成するノイズ除去手段と、該ノイズ除去手段によって作成されたネットワークの連結情報を元にネットワークを分節する分節手段と、該分節手段によって分節された結果を出力する出力手段と、を備えることを特徴とするネットワークの自動分節装置。

【請求項3】 コンピュータを、記号系列の構造を示すネットワークの自動分節装置であって、ネットワーク構造を読み込む読込手段と、該読込手段によって読み込まれたネットワーク構造から各ノードの連結情報を抽出する連結情報抽出手段と、該連結情報抽出手段によって抽出された各ノードの連結情報を元にネットワークを分節する分節手段と、該分節手段によって分節された結果を出力する出力手段と、を備えるネットワークの自動分節装置として機能させるためのプログラムを記録したことを特徴とするコンピュータ読み取り可能な記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は、テキストや音声や画像などを記号系列で表現する場合に、その記号系列の構造を示すネットワークの特徴を抽出することができるネットワークの自動分節装置及び自動分節プログラムを記録した記録媒体に関する。

【0002】

【従来の技術】 記号系列のデータのモデル化やその系列の特徴を表現するために、従来多くの表示方法が考えられている。その一つの方法として、特開平10-198391号公報に示されるネットワーク作成装置及び方法がある。この方法は、記号（シンボル）系列の構造を自動的に構造化し、ネットワークで表現する。すなわち、図9に示すように、このネットワークは、ノード1と、ノード1同士を結ぶアーク2とから構成される。ノード1はネットワークのある状態を表す有限の集合である。

アークはリンクとも呼ばれ、ある条件による遷移を表す。このため、アークはその向きに意味がある。アークのうち、アークが出ているノード1自身に戻るアークを自己ループアークと呼ぶ。自己ループアーク3は同じ入力が続く場合の状態の滞留を示す。

【0003】

【発明が解決しようとする課題】 図9のようなネットワーク構造は、ノード1、アーク2が少ない場合には、ネットワーク構造の特徴を人間が抽出することは可能である。しかしネットワーク構造が複雑で、ノード1、アーク2がともに多い場合には人が抽出することは困難であり、實際上すべての特徴を抽出することはできないという問題があった。

【0004】 例えば、図1に示すようなネットワーク構造において、Node i からNode i + 3 までの区間が「ABC」という記号列で表現されたとすると、ネットワーク全体では記号「A」は必ず記号「B」と、記号「B」は必ず記号「C」と接続するが、従来はネットワーク構造全体からこのような特徴的な部分を自動的に抽出する方法はなく、人間の視察に頼るしかなかった。

【0005】 大規模なネットワーク構造から元の記号系列データの連結情報を獲得することは、ネットワーク構造の特徴を把握する上で有効な手段ではある。しかしながら、従来手法ではネットワーク構造の連結情報を自動的に獲得する方法がなかった。

【0006】 本発明は、このような問題に鑑みてなされたものであって、その目的とするところは、記号系列の構造を示すネットワークから特徴のある接続構造を自動抽出することができるネットワークの自動分節装置及び自動分節プログラムを記録した記録媒体を提供することである。

【0007】

【課題を解決するための手段】 本発明は、記号系列の構造を示すネットワークの自動分節装置であって、ネットワーク構造を読み込む読込手段と、該読込手段によって読み込まれたネットワーク構造から各ノードの連結情報を抽出する連結情報抽出手段と、該連結情報抽出手段によって抽出された各ノードの連結情報を元にネットワークを分節する分節手段と、該分節手段によって分節された結果を出力する出力手段と、を備えるものである。

【0008】 また、本発明は、記号系列の構造を示すネットワークの自動分節装置であって、ネットワーク構造を読み込む読込手段と、該読込手段によって読み込まれたネットワーク構造から各ノードの連結情報を抽出する連結情報抽出手段と、該連結情報抽出手段によって抽出された各ノードの連結情報から所定のしきい値よりも頻度が少ないアークを除き、これにより入力アークが存在しなくなったノードを除いたネットワークを作成するノイズ除去手段と、該ノイズ除去手段によって作成されたネットワークの連結情報を元にネットワークを分節する

分節手段と、該分節手段によって分節された結果を出力する出力手段と、を備えるものである。これにより、ネットワーク構造を生成する際に、若干のノイズが加わった場合でも、ネットワーク構造から冗長なノード、アークを除去し、ネットワークの構造を自動的に分節することができる。

【0009】さらに、本発明は、コンピュータを、記号系列の構造を示すネットワークの自動分節装置であって、ネットワーク構造を読み込む読込手段と、該読込手段によって読み込まれたネットワーク構造から各ノードの連結情報を抽出する連結情報抽出手段と、該連結情報抽出手段によって抽出された各ノードの連結情報を元にネットワークを分節する分節手段と、該分節手段によって分節された結果を出力する出力手段と、を備えるネットワークの自動分節装置として機能させるためのプログラムを記録したコンピュータ読み取り可能な記録媒体である。

【0010】

【発明の実施の形態】以下、添付図面を参照しながら本発明の好適な実施の形態について詳細に説明する。図1は、本発明のネットワークの自動分節装置の原理を説明する図であって、ネットワーク構造を自動的に分節し、ネットワーク構造中の特徴のある部分を抽出できるようにしたものである。ネットワーク構造4は、ノード1とアーク2を用いて構成される。アーク2には1つのシンボルラベルが対応づけられ、一組のノードペア間には1つのアーク2が対応づけられている。同一の入力ラベルに対して自己ループアーク3の生成を許すようになっている。それぞれのノード1はいくつのアーク2を持つか、どのノード1に遷移するか、その時の入力ラベル情報、通過頻度等の情報を持つ。図1は、このようにして生成したネットワーク構造の一部を示している。

【0011】図1中のノードNode i へのアークは複数であるが、Node i からのアーク2は1本である。このNode i のようなノードを分節の始点とする。同様にNode i + 3へのアークは1本であるが、Node i + 3からのアークは複数存在する。このNode i + 3のようなノードを分節の終点とする。

【0012】図2は、ネットワーク作成時のノイズ情報による誤分節を防ぐために、ノード1及びアーク2のうち使用頻度の少ないものを除去する原理を示す。図2(a)において、Node(a)からの2本のアークの使用頻度を n_i 、 n_j とするとき、 n_i が n_j に比べて十分に小さい場合、Node(d)へのアークを取り除く。Node(d)へのアークが1本だけであるなら、Node(d)自身も削除し、図2(b)に示すようなネットワーク構造とする。

【0013】図3は、本発明の処理の流れを示すフロー図である。図3(a)において、時系列の情報を表すネットワーク構造を読み込み(ステップS30)、そのネットワーク構造から各ノードの連結情報を抽出し(ステッ

プS31)、そのネットワーク構造から頻度の少ないノード、アークを除いたネットワークを作成した後(ステップS32)、ネットワークを分節し(ステップS33)、分節が終了したか否かを判断し(ステップS34)、分節が終了したときその分節結果を出力し表示する(ステップS35)。

【0014】図4は、上記ネットワークの自動分節装置のシステム構成を示している。図4において、CPU100は内蔵のシステムプログラムに従ってシステム制御を行うほか、自動分節プログラムを実行してワークメモリ101上に分節結果を作成する。キーボード102からはネットワーク分節の起動の指示や分節を行うネットワークを入力する。ディスプレイ装置103はキーボード102から入力された情報やCPU100の演算処理結果を表示する。CD-ROMドライバ104は自動分節プログラムを記録したCD-ROM200の読み取り装置である。

【0015】図5は、本発明のネットワークの自動分節装置の第1実施の形態の構成を示している。このネットワークの自動分節装置は、ネットワーク構造を読み込むネットワーク情報入力装置10と、連結情報を抽出する連結情報抽出装置11と、ネットワークを分節する分節装置12と、それぞれの結果を表示する出力装置13から構成される。

【0016】図6は、第1実施の形態のネットワークの自動分節装置の処理の流れを示すフロー図である。記号系列の情報を表すネットワーク構造をネットワーク情報入力装置10で読み込み(ステップS60)、このネットワーク構造から連結情報抽出装置11で各ノードの連結情報を抽出し(ステップS61)、この情報を元に分節装置12でネットワークを分節し(ステップS62)、分節が終了したか否かを判断し(ステップS62)、分節が終了したときその分節した結果を出力し出力装置13に表示する(ステップS64)。前記ネットワーク情報入力装置10で読み込まれたネットワーク構造に対し、分節の始末端を抽出する。図7は、その詳細な流れを示すフロー図である。

【0017】図7において、変数初期化した(ステップS70)後、最初のノードについて、アークの数が1本(ステップS71)か、または2本かつそのうち一本は自己ループアークである場合(ステップS72)には、分節区間の長さLをインクリメントし(ステップS74)、次のノードについて同じ処理を繰り返す。それ以外の場合で、長さLが3以上の場合(ステップS75)には分節数nをインクリメントし(ステップS76)、長さLを初期化して、次のノードの処理を行う。長さLが3未満の場合には長さLを初期化し(ステップS77)、次のノードの処理を行う。すべてのノードについて処理を行った場合、終了する。また図3に戻り、終了の場合にはすべての分節結果を出力する。ここで、「ア

ークの数が1本か、または2本かつ1本が自己ループアーク」の状態の連続性により、分節の始末端を抽出できる。また、始端の判断は「ネットワーク構造からあるノードへのアークの数を計測する」ことにより判断できる。同様に終端の判断は「ネットワーク構造中のあるノードからのアークの数を計測する」ことにより判断できる。

【0018】図8は、本発明のネットワーク自動分節装置の第2実施の形態の構成を示している。この第2実施の形態のネットワーク自動分節装置は、冗長なネットワーク構造を除去する機能を有するものである。すなわち、連結情報を抽出する連結情報抽出装置11の次段に、使用頻度の少ない部分を取り除くノイズ除去装置14を設け、このノイズ除去装置14及び分節装置12のそれぞれの装置での処理結果を出力装置13に入力して表示する。

【0019】上記ノイズ除去装置14の処理を図2及び図3(b)のノイズ除去の処理の流れを示すフロー図を用いて説明する。図2のNode(a)からのアークの総頻度を nt 、あるアークの頻度を nj 、しきい値を x とすると、 $nj/nt < x$ ならば(ステップS38)、そのアークを削除する(ステップS39)。また、入力アークのなくなったノードも削除する。前述の処理をすべてのノード、アークに対して行い、新たなネットワーク構造を生成する。以下の処理は第1実施の形態と同様である。なお、本発明は上記実施の形態に限定されるものではない。

【0020】以上説明したネットワークの自動分節装置は、このネットワークの自動分節装置を機能させるためのプログラムでも実現される。このプログラムはコンピュータで読み取り可能な記録媒体に格納されている。本発明では、この記録媒体として、図4に示されているワークメモリ101そのものがプログラムメディアであってもよいし、また外部記憶装置としてCD-ROMドライブ104等のプログラム読み取り装置が設けられ、そこに記録媒体を挿入することで読み取り可能なCD-ROM200等のプログラムメディアであってもよい。いずれの場合においても、格納されているプログラムはCPU100がアクセスして実行させる構成であってもよいし、あるいはいずれの場合もプログラムを読み出し、読み出されたプログラムは、図示されていないプログラム記憶エリアにダウンロードされて、そのプログラムが実行される方式であってもよい。このダウンロード用のプログラムは予め本体装置に格納されているものとする。

【0021】ここで、上記プログラムメディアは、本体と分離可能に構成される記録媒体であり、磁気テープやカセットテープ等のテープ系、フロッピーディスクやハードディスク等の磁気ディスクやCD-ROM/MO/MD/DVD等の光ディスクのディスク系、ICカード

／光カード等のカード系、あるいはマスクROM、EPROM、EEPROM、フラッシュROM等による半導体メモリを含めた固定的にプログラムを担持する媒体であってもよい。

【0022】さらに、図示されていないが、外部の通信ネットワークとの接続が可能な装置を備えている場合には、その通信接続装置を介して通信ネットワークからプログラムをダウンロードするように、流動的にプログラムを担持する媒体であってもよい。なお、このように通信ネットワークからプログラムをダウンロードする場合には、そのダウンロード用プログラムは予め本体装置に格納しておくか、あるいは別な記録媒体からインストールされるものであってもよい。なお、記録媒体に格納されている内容としてはプログラムに限定されず、データであってもよい。また、ネットワークの読み込みは記号系列を読み込んでネットワークを形成するものでもよい。

【0023】

【発明の効果】以上、詳述したように、本発明のネットワークの自動分節装置及び自動分節プログラムを記録した記録媒体によれば、記号系列の構造を示すネットワークの接続構造を解析し、その特徴に基づいて分節することができる。この処理は初期状態から行われるため、ユーザは手動操作による初期化作業を行う必要がなく、完全自動で特徴のある部分を抽出することができる。

【図面の簡単な説明】

【図1】ネットワーク構造の分節を説明する図である。

【図2】ネットワーク構造からノイズによる冗長な構造を取り除く説明をする図である。

【図3】本発明の処理の流れを示すフロー図である。

【図4】本発明のネットワークの自動分節装置のシステム構成図である。

【図5】本発明のネットワークの自動分節装置の第1実施の形態の構成を示すブロック図である。

【図6】第1実施の形態の処理の流れを示すフロー図である。

【図7】第1実施の形態の分節の詳細な流れを示すフロー図である。

【図8】本発明のネットワークの自動分節装置の第2実施の形態の構成を示すブロック図である。

【図9】従来の一般的なネットワーク構造を示す図である。

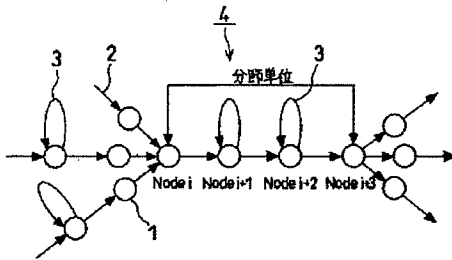
【符号の説明】

- 1 ノード
- 2 アーク
- 3 自己ループアーク
- 4 ネットワーク構造
- 10 ネットワーク情報入力装置
- 11 連結情報抽出装置
- 12 分節装置

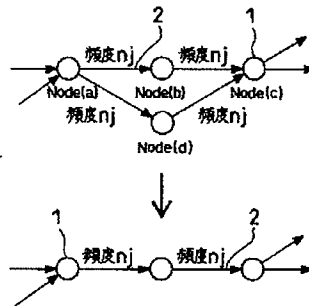
13 出力装置

* * 14 ノイズ除去装置

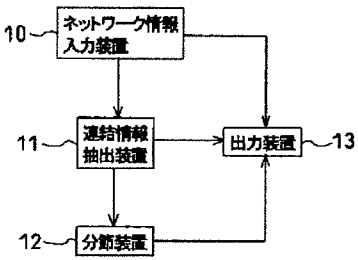
【図1】



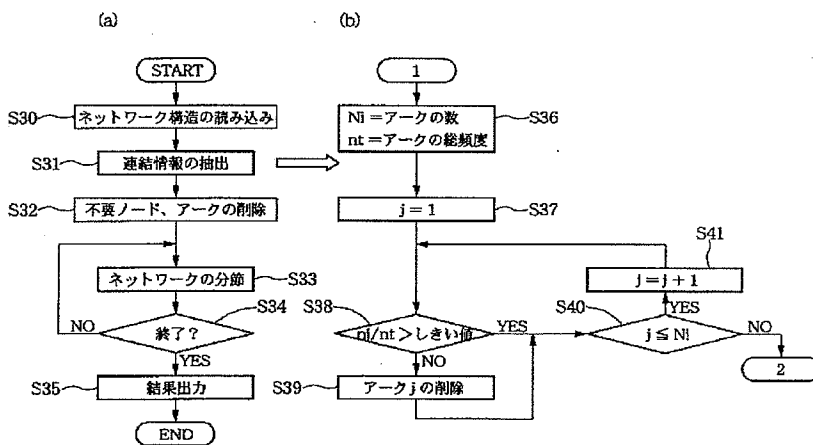
【図2】



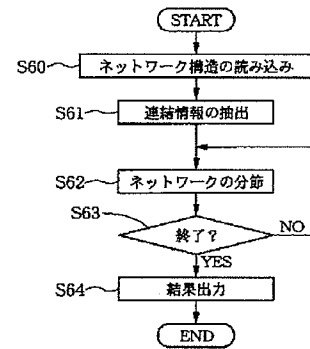
【図5】



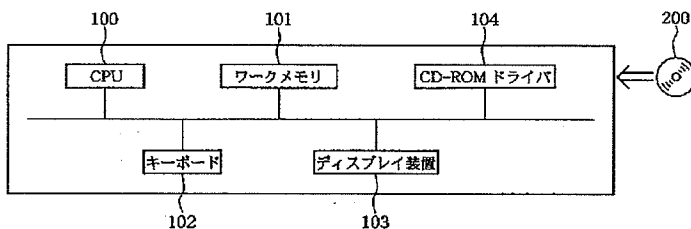
【図3】



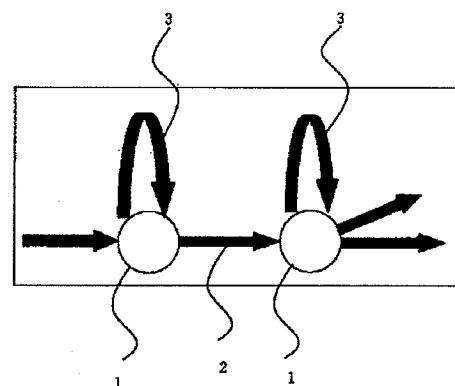
【図6】



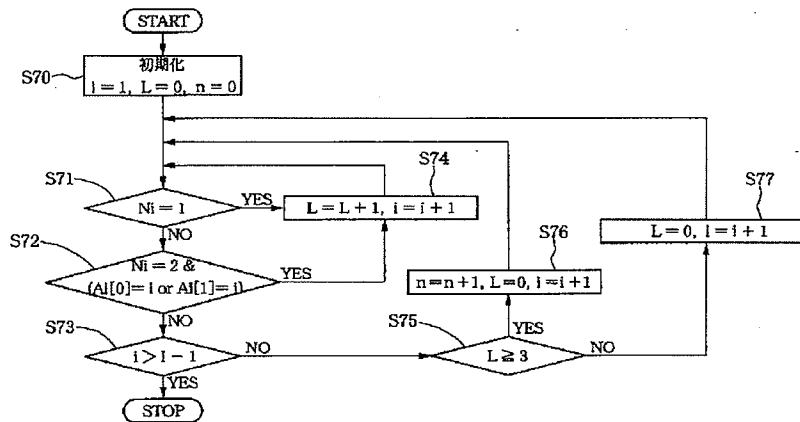
【図4】



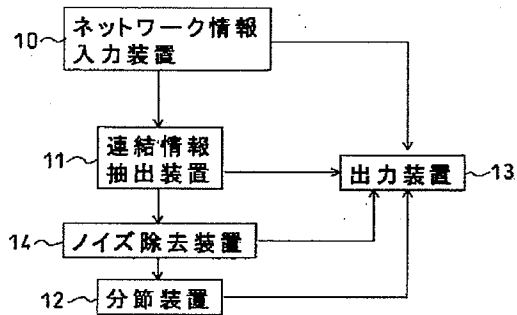
【図9】



【図7】



【図8】



フロントページの続き

(72)発明者 岡 隆一
茨城県つくば市竹園1-6-1 つくば三
井ビル 技術研究組合 新情報処理開発機
構 つくば研究センタ内

Fターム(参考) 5D015 GG00 HH03 HH23